Record Request RR-DTE-91

Q. (1) Please rerun these equations with throughput instead of sales as an independent variable. Please indicate the computer software used to estimate the regressions;

(2) Indicate the level of statistical significance of the estimates selected by the Company to determine whether or not an independent variable has explanatory power;

(3) Please state all assumptions and underlying assumptions of the methods used by the Company to estimate the equations, and explain how the Company tested these assumptions to ensure that no assumptions are violated. Please provide evidence to support your answer;

(4) Specify which test or tests, were performed in the revision analysis to detect serial autocorrelation in the residuals and multicollinearity in the data. Please provide evidence to support your answer. If applicable, please correct the serial autocorrelation problem and multicollinearity problem;

(5) Compute and present the correlation matrix among all independent variables using the regression equations, if more than one variable is used.

(6) Please provide the output files and all the data used for the Company to estimate the equation.

(7) To the extent that the Company is unable to respond fully to the Department's request, please provide a discussion of what the Company would propose to do should a more detailed analysis be undertaken by the Company in the future.

A. (1) Please see Attachment RR-DTE-91(a) and (b) submitted herewith. Attachment RR-DTE-91(a) provides a summary of the equations tested and Attachment RR-DTE-91(b) provides the results of the regression analysis. The Company used SAS/STAT statistical software developed by SAS Institute, Inc. to estimate the regression equations presented in the Excel spreadsheets, which is submitted herewith as the attachment.

(2) To determine whether an independent variable has explanatory power, the Company sought a 0.05 level of significance as indicated by a t-statistic of approximately 2 or greater on the estimated variable.

(3) In the development of the marginal cost regression analysis presented herewith, the Company used the assumptions listed below. The tests that the Company performed to verify the assumptions are described in Response 4 below.

?? Distribution plant is a function of either throughput and population density, or number of customers and population density,

?? Disaggregating throughput and numbers of customers into residential and commercial/industrial categories will enhance the explanatory power of the equations,

?? The relationship between distribution plant and the independent variables can be statistically determined and will give reasonable marginal costs estimates,

?? The relationship between distribution plant and throughput is positive,

?? The relationship between distribution plant and number of customers is positive,

?? The relationship between distribution plant and population density is negative,

?? The throughput data (in total and disaggregated) and population density data are not correlated,

?? The error terms of the equations are not serially correlated.

(4) The Company tested the assumptions listed above using the SAS statistical software to run multivariate linear and logarithmic equations. To test the statistical strength of each equation, the Company evaluated the R-square statistic seeking an R-square of .85 or better. As stated in Response 2, above, the Company tested the strength of individual variables at the 0.05 significance level. The t-statistics associated with each parameter are listed in Attachment (b), along with the level of significance under the heading "Pr<|t|".

The Company tested for serial autocorrelation in the residuals of each equation using the Durbin-Watson statistic calculated by the SAS software and comparing it to the table of Durbin-Watson statistics presented in Statistics for Economists**,** Ralph Beals, Rand McNally Publishing, Chicago (1972) Table A-5. The Durbin-Watson statistics calculated for each equation are labeled "DW" on Attachment RR-DTE-91(b). The upper and lower limits of the indeterminant range of Durbin-Watson statistics are also shown in the Attachment as "du" and "dl" respectively. A Durbin-Watson statistic below the lower limit indicates the presence of serial autocorreleation. A Durbin-Watson statistic above the upper limit indicates no serial autocorrelation. As the term implies, a Durbin-Watson statistic between the lower and upper value of the indeterminant range is indeterminant. The SAS statistic software has an autocorrelation correction routine that uses the Maximum Likelihood Estimator Method. This routine was run and the results of the corrected equation are presented for each equation. The corrected equations are presented on the page immediately following the page containing the original equation and are identified with the postscript "c" after the equation's case number.

The Company tested for multicollinearity by evaluating the correlation coefficient matrix for pairs of variables that showed high degree of correlation, generally greater than 75%. To correct for evidence of multicollinearity, the Company dropped one of the correlated variables and re-ran the regression. In most cases,

the population density variable was dropped because it correlated with one of the other variables and because it has a weaker theoretical basis for explaining distribution plant than either throughput or customers.

The corrections for serial autocorrelation and multicolinearity are presented in Attachment RR-DTE-91(b) on the pages immediately following the results of the original equations.

(5) The correlation matrix for each equation is presented on each page of Attachment RR-DTE-91(b) on or about line 37 under the heading "Correlation Estimate."

(6) The output and all of the data used by the Company to estimate the equations are presented in Attachment RR-DTE-91(b).

(7) If the Company were to pursue a more detailed statistical analysis of marginal costs, the Company would propose to do a statistical analysis -- to the extent possible -- on each of the cost components estimated in the marginal cost study presented as Exhibit KEDNE/ALS-2. These cost components are distribution capacity cost, capacity-related production expenses, capacity-related transmission and distribution expenses, customer-related operating expenses, customer accounting expenses, administrative and general loading factors, and miscellaneous loading factors. The analysis would specify an equation for each one of these cost components, using the cost component as the dependent variable and the relevant cost-causation factors as the independent variables. For example, customer-related expenses would regressed against number of customers.

The Company would also supplement the data already available and presented in the marginal cost study with additional data series to the extent that such data is available. The analysis would include the relevant statistical significance tests, such as those presented in this response, to determine the explanatory strength of each equation. Lastly, the results would be analyzed to evaluate the appropriateness of the marginal cost estimates. The entire process would be documented to include a description of the a priori assumptions, the equation specifications, the statistical tests and corrections made to the equations, an evaluation of the results and a presentation of the final equations selections, along with an explanation was why the final equations were chosen. The Company would anticipate that the full study, including the documentation, would take approximately three to six months to complete.